

Result No.	Query Match	Score	Length	ID	Description	Pred. No.
1	98%	0.76	100	1	100% match	1
2	98%	0.76	100	2	100% match	2
3	98%	0.76	100	3	100% match	3
4	98%	0.76	100	4	100% match	4
5	98%	0.76	100	5	100% match	5
6	98%	0.76	100	6	100% match	6
7	98%	0.76	100	7	100% match	7
8	98%	0.76	100	8	100% match	8
9	98%	0.76	100	9	100% match	9
10	98%	0.76	100	10	100% match	10
11	98%	0.76	100	11	100% match	11
12	98%	0.76	100	12	100% match	12
13	98%	0.76	100	13	100% match	13
14	98%	0.76	100	14	100% match	14
15	98%	0.76	100	15	100% match	15
16	98%	0.76	100	16	100% match	16
17	98%	0.76	100	17	100% match	17
18	98%	0.76	100	18	100% match	18
19	98%	0.76	100	19	100% match	19
20	98%	0.76	100	20	100% match	20
21	98%	0.76	100	21	100% match	21
22	98%	0.76	100	22	100% match	22
23	98%	0.76	100	23	100% match	23
24	98%	0.76	100	24	100% match	24
25	98%	0.76	100	25	100% match	25
26	98%	0.76	100	26	100% match	26
27	98%	0.76	100	27	100% match	27
28	98%	0.76	100	28	100% match	28
29	98%	0.76	100	29	100% match	29
30	98%	0.76	100	30	100% match	30
31	98%	0.76	100	31	100% match	31
32	98%	0.76	100	32	100% match	32
33	98%	0.76	100	33	100% match	33
34	98%	0.76	100	34	100% match	34
35	98%	0.76	100	35	100% match	35
36	98%	0.76	100	36	100% match	36
37	98%	0.76	100	37	100% match	37
38	98%	0.76	100	38	100% match	38
39	98%	0.76	100	39	100% match	39
40	98%	0.76	100	40	100% match	40
41	98%	0.76	100	41	100% match	41
42	98%	0.76	100	42	100% match	42
43	98%	0.76	100	43	100% match	43
44	98%	0.76	100	44	100% match	44
45	98%	0.76	100	45	100% match	45
46	98%	0.76	100	46	100% match	46
47	98%	0.76	100	47	100% match	47
48	98%	0.76	100	48	100% match	48
49	98%	0.76	100	49	100% match	49
50	98%	0.76	100	50	100% match	50
51	98%	0.76	100	51	100% match	51
52	98%	0.76	100	52	100% match	52
53	98%	0.76	100	53	100% match	53
54	98%	0.76	100	54	100% match	54
55	98%	0.76	100	55	100% match	55
56	98%	0.76	100	56	100% match	56
57	98%	0.76	100	57	100% match	57
58	98%	0.76	100	58	100% match	58
59	98%	0.76	100	59	100% match	59
60	98%	0.76	100	60	100% match	60
61	98%	0.76	100	61	100% match	61
62	98%	0.76	100	62	100% match	62
63	98%	0.76	100	63	100% match	63
64	98%	0.76	100	64	100% match	64
65	98%	0.76	100	65	100% match	65
66	98%	0.76	100	66	100% match	66
67	98%	0.76	100			

```

FEATURES
source      Biology, Pasadena, CA 91101, USA
            Location/Qualifiers
            1..2399
            /organism="Rattus norvegicus"
            /strain="Sprague Dawley"
            /clone="E13.5"
            279..1070
            /gene="DRG11"
            /codon_start=1
            /product="paired-like homeodomain transcription factor"
            /db_xref="PID:g1144015"
            /translation="MFIFHCPQLEGTAFGNHSTGDFDLRRKORNRRTFALQO
LEALEAVFATHYPDVTFREELAMKINLEARVQVWFQNRRAKRWTERGASQDEPGA
KEPAEYTPPPVNRPPGQDQARKEALEAQQSLGRVTPAGPFPSCLPPTLIN
TATVAQLSHVSLKGGPLSCCVPDPMLGSLFTYGCQSNRTASVAALRWKAREHSE
AVLASANLLPSTSSSPSPASKQVPPESQDKPSPTKEQSEKSV"
            - polyA_site
            2399
BASE COUNT 511 a 681 c 658 g 549 t
ORIGIN
Jury Match 99.0%; Score 2399; DB 88; Length 2399;
Best Local Similarity 100.0%; Pred. No. 0.00e+00;
Matches 2399; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1 9cagaggtaggcagggttcccgagcgctctcccggtccctgctctggtggtgggct 60
Qy 1 GCAGAGGTAGGAGGTTCCGAGCGCGTCTCCGCGTCCCTGCTCTGGGGCTTGGGGCT 60
Db 61 ccacggcttcttggcccgagctgctgcgctgcagatggccttgcgcatgcgcggacc 120
Qy 61 CCACGGGCTTCTGGCCCGAGCTGCTGGCGGTGCAGATGGCTTGGCGGATCGCGGACC 120
Db 121 ccgctgcggtgccaagtcaggcttgcgtggtgagacccctggaaccagagcagac 180
Qy 121 CCCTGCGGTGCGCAAGTGCAGGCGCTTGTGCTGGGACCCCTGAGAACCCAGAGCCAGAC 180
Db 181 tgtgtcagcttgcagcgccgagcgcacgcgcgcgcacaaagtctgtcagcgcccatcag 240
Qy 181 TGTGCTCAGCTTGCAGCGCGGAGCCAGCCAGCGGCACAAAGTCTGTGAGCGCCCATCAG 240
Db 241 tcttggtccagcgtcagggcccatccgacccgtcggtggtggtttttatttccactgcgc 300
Qy 241 TCCTTGTGTCAGCGCTCAGGGCGCCATCGGACGCTCGGGGATGTTTTATTTCACCTGCCCGC 300
Db 301 cacagctagagggcacagcgcttctgttaaccactcaggggatttgcagtagtggt 360
Qy 301 CACAGCTAGAGGGCACAGCGCTTTTGTAAACCACTCTACGGGGGATTTGTATGATGGGT 360
Db 361 ttcttagaagaaacagcgagaaatcggaacacttcgctcttcagcaggttgaagctc 420
Qy 361 TTCTTAGAAGAAACAGCGCAGAAATCGGACAACTTCGCTCTTCACGAGTTGGAAGCTC 420
Db 421 tggaggcagtttggcccaaacacactaccagatgtcttcacagagaagctagcca 480
Qy 421 TGGAGGAGGCTTTTGCCCAACACACTACCCAGATGTCTTCACAGAGAAGAGCTAGCCA 480
Db 481 tgaataaaacctcacagaagccagagtcaggttgggttcagaaaccgagagcagaagt 540
Qy 481 TGAATAAAACCTCACAGAAGCCAGAGTGCAGGTTTGGTTCAGAAACCCGAGAGCCAAAGT 540
Db 541 ggaaggaagacagagagggggtccttcacaggaacacaggggcttaaggaacccatggcag 600
Qy 541 GGAGGAAGACAGAGAGAGGGGCTCTGACCAAGNACCAGGGGCTAAGGAACCCATGGCAG 600
Db 601 aggtgacacacccccagtgagaaacatcaactctccacccccaggggagaccagggccggg 660
Qy 601 AGGTGACACACCCCCAGTGAGGAACATCAACTCTCCACCCCCAGGGGACACAGGCCCGGG 660
Db 661 gcaagaagagggccctgagggcccgacagagcctggagcagcagtggtggcccgccgggc 720
Qy 661 GCAAGAAGAGGGCCCTGGAGGCCCGACAGAGCCTGGGACGCACAGTGGGGCCCGCGGGC 720

```

```

|||||
QY 1801 GTGTTCACTTGCAGGCCGCTCTCTTCAGAGAGATGATGGAGTTCACCAATCAGAAAT 1860
Db 1861 gtagcttttagccaggaagacagagctcttgaagcgttagaagaaatccctaggaag 1920
QY 1861 GTAGCTTTGTAGCCAGGAAAGACACAGAGTCTCTGAGCGGTAGGAATCCTTAGAAG 1920
Db 1921 gcccttaataactatgcccagatgagctgccctcttctctatccccctatgtcgagag 1980
QY -1921 GCCCTTAATAACTATTATGCCACATGAGTGCCTCTCTTCTATCCCTATGTCGAGAG 1980
Db 1981 gttgacagacagaaagccaggaagatgaactccgtgtgagcagaagaaatgagtcacaa 2040
QY 1981 GTTGACGACAGAGAAAGCCAGGAAGATGACTCCGTGTGGCAGAGAGAAATGGAGTCCAA 2040
Db 2041 agggccaactttcacagagattctccagcttagcgtggtgtgtctcttcacagcga 2100
QY 2041 AGGGCCAACCTTTCACAGAGATTCTCCCGCAGTTTACGTGGCTGTGTCTTCACGCGA 2100
Db 2101 tggtagcttcggagagatcagagggagatgtgcaatagcatgagccccctcctggcccg 2160
QY 2101 TGGTAGCTTCGGAGAGATCAGAGGAGATGTGCAATAGCATGAGCCCCGCTCTCTGGCCCG 2160
Db 2161 ggtcctggaaagtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt 2220
QY 2161 GGTCTCTGGAAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2220
Db 2221 atattcttcagtggtctctcatttcagcttctgataatcagagcgtgtgtctccc 2280
QY 2221 ATATTGCTTCATGATGCTGTGTCTCATTTACGTTCTGTATATCGAGAGCGTGTGTCTCC 2280
Db 2281 aggcgtctgctcccttaactcttcagagcagccccctcctgtataactcctaagt 2340
QY 2281 AGCGCTCTGCTCCCTTAACCTTTCAGGAGCAGCCCCCTCCCTCTGTAATCTCTTAAGT 2340
Db 2341 gtagctgctcactacgttactcaacacacattgtagtgaacacactgtctctctta 2399
QY 2341 GTATCGTCTCACTTACGCTTACTGCAACATTTGATGGAGACACACTGTCTCTCTTA 2399

RESULT 2 HUMPHOXIA 1433 bp mRNA PRI 07-JAN-1995
LOCUS Human homeobox protein (PHOX1) mRNA, 3' end.
ACCESSION M95929
NID q189946
KEYWORDS homeobox protein.
SOURCE Homo sapiens cDNA to mRNA.
ORGANISM Homo sapiens
REFERENCE Gruenberg,D.A., Natesan,S., Alexandre,C. and Gilman,M.Z.
AUTHORS Human and Drosophila homeodomain proteins that enhance the DNA
TITLE binding activity of serum response factor
JOURNAL Unpublished (1992)
FEATURES
source
Location/Qualifiers
1..1433
/organism="Homo sapiens"
/cell_line="U118-MG"
/cell_type="gliblastoma"
<1..600
/gene="PHOX1"
/codon_start=1
/product="homeobox protein"
/db_xref="PID:g189947"
/translaton="LDSPNDLTLQAKNFVSHLLDEEAGMVAQAQDENVGAGR
SLESPLTSGSDTPOQNDLNSEKKRQRNRRTFNSSQLALERVFFTHYPD
AFVREDLARNVLTARVOVTFQNRRAKRRNRRLANKNLKLSYSGDVTAVEQP
IVRPAPRPTDYLWGTASPYRSSSLPRCLLEHNGF"
misc_feature
233..405
/gene="PHOX1"
/standard_name="homeodomain"
BASE COUNT 391 a 373 c 364 g 305 t

```

```

ORIGIN
Query Match 3.4%; Score 82; DB 78; Length 1433;
Best Local Similarity 71.6%; Pred. No. 1.07e-22;
Matches 136; Conservative 0; Mismatches 54; Indels 0; Gaps 0;

Db 220 agaaagcgcgaagaatagacaacattcaatagcagccagcttggagcgt 279
|||||
QY 369 AGAAACACGCGCAGAAATCGGACACCTTCCTCTTCAGCAGTTTGGAGCTCTGGAGCA 428
|||||
Db 280 gctcttgagcggacacactatcctgagctcttggcagagacacctgcccgcgggtg 339
|||||
QY 429 GTCCTTGCCAAACACATACCCAGATGCTTCCACGAGAGAGAGTAGCCATGAAATA 488
|||||
Db 340 aacctcacgagggagagtgagtggtgttcagaaacgaagcgaagtccgcagg 399
|||||
QY 489 AACCTCACAGAGCGAGAGTGCAGTGTGTTTCCAGAACCCGAAGACCAAGTGGAGGAG 548
|||||
Db 400 aatgagagag 409
QY 549 ACAGAGAGAG 558

RESULT 3 MUSCHX10A 3089 bp mRNA ROD 12-JAN-1995
LOCUS Mus musculus (clone 6) homeobox protein (Chx10) mRNA, complete cds.
ACCESSION L34808
NID G623165
KEYWORDS homeobox protein.
SOURCE Mus musculus (strain CD1/129sv) cDNA to mRNA.
ORGANISM Mus musculus
REFERENCE Liu,I.S., Chen,J.D., Ploder,L., Vldgen,D., van der Kooy,D.,
AUTHORS Ramnus,V.I. and McInnes,R.R.
TITLE Developmental expression of a novel murine homeobox gene (Chx10):
evidence for roles in determination of the neuroretina and inner
nuclear layer
JOURNAL Neuron 13 (2): 377-393 (1994)
MEDLINE 94338694
FEATURES
source
Location/Qualifiers
1..3089
/organism="Mus musculus"
/strain="CD1/129sv"
/map="chromosome 12"
1..1086
/gene="Chx10"
/codon_start=1
/product="homeobox protein"
/db_xref="PID:g623166"
/translaton="MTGKAGALSKPKSETVAKSTSGGAPARCFTGFIQIETLGNKEP
PSSHPRAALDGLAPGHLAARSLSBAGVSGMLGPGGLPGFVTOPTLEVLSDPOS
VHLQPLRASGLDTQSATSDSDVSSDRKMSKALNQTKKRRHRRTFTSYQL
EELKAFNEAHPDYVAREMLAMKTELPEDRIQVFNRRRAKRRKRCWRSSVMAE
YGLYKAMVHSIPLPESTLKAKGIMDSAPLWLGHHKHSLEAAASGRKPEVROA
LPKLDKMEBERAPEAQAAISQELRENSIAALRAKAEHSTKVLGTVSGPDSLARNA
EKPEEDATEEDRPAEKLSPPQLEDA"
polyA_site 3089
BASE COUNT 703 a 897 c 800 g 689 t
ORIGIN
Query Match 3.3%; Score 80; DB 84; Length 3089;
Best Local Similarity 72.2%; Pred. No. 1.02e-21;
Matches 130; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Db 443 agcggcgacacaggaacattcttactctaccagcagctagagctggagaagattca 502
|||||
QY 376 AGCGCAGAAATCGGACAACTTCGCTCTTCAGCAGTGTGGAAGCTCTGGAGGAGTCTT 435
|||||
Db 503 atgaagcccaatccacagatgtctacgcccgggagatgtggtccatgaaacggagctcc 562
|||||

```





## ORIGIN

```

QY 429 GTCTTGGCCCAACACACTACCCAGATGCTCTTCCACAGAGAGAGAGTACCCATGAAATA 488
Db 585 gagctccagagacagaatacaggtatgtttcagaacccgctgctgaagtgagagaa 644
QY 489 AACCTCAGAGAGCCAGAGTGCAGGTTTGGTTCCAGAACCGAGAGCCCAAGTGGAGGA 548
Db 645 agggaga 651
QY 549 ACAGAGA 555

RESULT 11
LOCUS MALLX3 1744 bp RNA 30-MAR-1995
DEFINITION M.auratus mRNA for alx3.
ACCESSION X81403
NID g587455
KEYWORDS alx3 gene.
SOURCE golden hamster.
ORGANISM Mesocricetus auratus
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Eutheria; Rodentia; Sciurognathi; Myomorpha; Muridae;
Cricetinae; Mesocricetus.
REFERENCE 1 (bases 1 to 1744)
AUTHORS German,M.S.
TITLE Direct Submission
JOURNAL Submitted (05-SEP-1994) M.S. German, HORMONE RESEARCH INST.,
UNIVERSITY OF CALIFORNIA SAN FRANCISCO, HSW 1090, BOX 0534, SAN
FRANCISCO, CA 94143-0534, USA
REFERENCE 2 (bases 1 to 1744)
AUTHORS Rudnick,A., Ling,T.Y., Odagiri,H., Rutter,W.J. and German,M.S.
TITLE Pancreatic beta cells express a diverse set of homeobox genes
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 91 (25), 12203-12207 (1994)
MEDLINE 95083670
FEATURES
Source Location/Qualifiers
1..1744
/organism="Mesocricetus auratus"
/tissue_type="pancreatic islet"
/cell_type="beta-cell"
/cell_line="HIT-T15.M2.2.2"
/clone_lib="MG 1343.1"
/clone_lib="lambda gt11"
21..1061
/gene="alx3"
/codon_start=1
/db_xref="PID:g587456"
/translation="MAPDPERCAPFVGPAGPAAAGDPPPGQSTDSAPHLHPA
PPRPLRSFPCGPLEYLPAPKPKYLDLGPAPVINGHFYEGPAEEKASK
AASFQPLVDCRGPRGSPVQSGFGLASLRVPLSPGLPDMELAKSKRRNR
TTFSPQLEELKVEFKTHYPDVYAREQLALTDITEARVQVFNRRARWRKRYG
KMQEGRFTTAYDLSVLPRTDHPQLNSLWPGSGSPGCLMSPEGIPSCMSP
YSHGNVAGFMGVPAAPAHPIYSIHGFPALGHSFEPSPDGYKSPSLSVRYK
PREPPLNLNWT"
BASE COUNT 365 a 559 c 472 g 348 t
ORIGIN
Query Match 2.9%; Score 70; DB 81; Length 1744;
Best Local Similarity 69.4%; Pred. No. 6.65e-17;
Matches 125; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

Db 480 agcaagaagcgtcggaacccagcaacattccagcacattccagtggaagagtgagaa 539
QY 369 AGAAAACAGCGCAGGAATCGGACACCTTCGCTCTTCAGCAGTGTGAAGCTCTGGAGCA 428
Db 540 gtcttcagaaaccactaccgcagctgtagtcccgagagcagctggttaccagcag 599
QY 429 GTCTTGGCCCAACACACTACCCAGATGCTCTTCCACAGAGAGAGTACCCATGAAATA 488
Db 600 gacctgacagagcccggttacagctgctgttccagaacccagcagccagtgccggaag 659
QY 489 AACCTCAGAGAGCCAGAGTGCAGGTTTGGTTCCAGAACCGAGAGCCCAAGTGGAGGA 548

RESULT 12

```

```

LOCUS GPRX2 849 bp RNA 29-MAR-1996
DEFINITION G.gallus Prx-2 (S8) mRNA.
ACCESSION X79695
NID g558376
KEYWORDS Prx-2 gene.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
Vertebrata; Archosauria; Aves; Neognathae; Galliformes;
Phasianidae; Phasianinae; Gallus.
REFERENCE 1 (bases 1 to 849)
AUTHORS Leussink,B., Brouwer,A., el Khattabi,M., Poelmann,R.E.,
Gittenberger-de Groot,A.C. and Meijlink,F.
TITLE Expression patterns of the paired-related homeobox genes MHox/Prx1
and S8/Prx2 suggest roles in development of the heart and the
forebrain
JOURNAL Mech. Dev. 52 (1), 51-64 (1995)
MEDLINE 96076127
REFERENCE 2 (bases 1 to 849)
AUTHORS Meijlink,F.
TITLE Direct Submission
JOURNAL Submitted (10-JUN-1994) F. Meijlink, Netherlands Inst. for
Developmental, Biology, Hubrecht Laboratory, Uppsalalaan 8, 3584 CT
Utrecht, NETHERLANDS
FEATURES
Source Location/Qualifiers
1..849
/organism="Gallus gallus"
/dev_stage="10 days"
/clone_lib="lambda gt11 chicken embryo's libraries one from
Clontech and one from Karolinska Institute"
<1..498
/gene="Prx-2 (S8)"
/codon_start=1
/db_xref="PID:g558377"
/translation="ESLSPRGVAKRKKKORRNFTTNSOQLALERVFERHYPDFAF
VREELARVNLSEARVQVFNRRARFRNRERAMLRASLLKSYSOEAAIEQPMAP
RPTALSPEYLSWSSSSPYSTVPSYSSGTATAGGVNMANSIASLRKAKEFSLHQNO
VPTVN"
BASE COUNT 208 a 266 c 222 g 153 t
ORIGIN
Query Match 2.8%; Score 69; DB 50; Length 849;
Best Local Similarity 70.4%; Pred. No. 1.98e-16;
Matches 119; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Db 43 aaacagagagaaacccagcacattccagcagccagctcagctcgagaggttc 102
QY 372 AACAGCGCAGAAATCGGACACCTTCCTCTTCAGCAGTGTGAAGCTCTGGAGCAGTC 431
Db 103 tttagcgggacacattaccccgagccttcgtcgggaggagctgcccagaggggtgaac 162
QY 432 TTGCCCCAACACACTACCCAGATGCTCTTCACAGAGAGAGTACCCATGAAATAAC 491
Db 163 ctacgagggcagagtcagctggttggtttcagaacccaggggccaagt 211
QY 492 CTCACAGAAGCCAGAGTGCAGGTTTGGTTCCAGAACCCGAGAGCCCAAGT 540

RESULT 13
LOCUS S69088 895 bp mRNA 22-SEP-1994
DEFINITION gMHox-homeobox (chickens, stage 17 embryo, heart, mRNA, 895 nt).
ACCESSION S69088
NID g545359
KEYWORDS chickens heart stage 17 embryo.
SOURCE Gallus sp.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 895)
AUTHORS Kuratani,S., Martin,J.F., Wawersik,S., Lilly,B., Eichele,G. and
Olson,E.N.
TITLE The expression pattern of the chick homeobox gene gMHox suggests a
role in patterning of the limbs and face and in
compartmentalization of somites

```









\*\*\*\*\*



(TM)

\*\*\*\*\*

Release 2.1d John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by Intelligenetics, Inc.

MPsrch\_nn n.a. - n.a. database search, using Smith-Waterman algorithm

'in on: Fri May 30 16:06:24 1997; MasPar time 1700.26 Seconds  
912.837 Million cell updates/sec  
-dbular output not generated.

Title: >US-08-701-278-1  
Description: (1-2424) from US08701278.seq  
Perfect Score: 2424  
N.A. Sequence: 1 GCAGAGGTAGGACGGGTTC...GAAAAA...TTTTTTTTTTTTTT  
Comp: CGTCTCCATCCGTCACGAGG

Scoring table: TABLE default  
Gap 6

Nmatch STD : Dbase 0; Query 0

Searched: 886179 seqs, 320143548 bases x 2

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: EST-STS  
1:EST1 2:EST2 3:EST3 4:EST4 5:EST5 6:EST6 7:EST7 8:EST8  
9:EST9 10:EST10 11:EST11 12:EST12 13:EST13 14:EST14  
15:EST15 16:EST16 17:EST17 18:EST18 19:EST19 20:EST20  
21:EST21 22:EST22 23:EST23 24:EST24 25:EST25 26:EST26  
27:EST27 28:EST28 29:EST29 30:EST30 31:EST31 32:EST32  
33:EST33 34:EST34 35:EST35 36:EST36 37:EST37 38:EST38  
39:EST39 40:EST40 41:EST41 42:EST42 43:EST43 44:EST44  
45:EST45 46:EST46 47:EST47 48:EST48 49:EST49 50:EST50  
51:EST51 52:EST52 53:EST53 54:EST54 55:EST55 56:EST56  
57:EST57 58:EST58 59:EST59 60:EST60 61:EST61 62:EST62  
63:EST63 64:EST64 65:EST65 66:EST66 67:EST67 68:EST68  
69:EST69 70:EST70 71:EST71 72:EST72 73:EST73 74:EST74  
75:EST75 76:EST76 77:EST77 78:EST78 79:EST79 80:EST80  
81:EST81 82:EST82 83:EST83 84:EST84 85:EST85 86:EST86  
87:EST87 88:EST88 89:EST89 90:EST90 91:EST91 92:EST92  
93:EST93 94:EST94 95:EST95 96:EST96 97:EST97 98:EST98  
99:EST99  
EST-STS-TWO  
100:EST100 101:EST101 102:EST102 103:EST103 104:EST104  
105:EST105 106:EST106 107:EST107 108:EST108 109:EST109  
110:EST110 111:EST111 112:EST112 113:EST113 114:EST114  
115:EST115 116:EST116 117:EST117 118:EST118 119:EST119  
120:EST120 121:EST121 122:EST122 123:EST123 124:EST124  
125:EST125 126:EST126 127:EST127 128:EST128 129:EST129  
130:EST130 131:EST131 132:EST132 133:EST133 134:EST134  
135:EST135 136:EST136 137:EST137 138:EST138 139:EST139  
140:EST140 141:EST141 142:EST142 143:EST143 144:EST144  
145:EST145 146:EST146 147:EST147 148:EST148 149:EST149  
150:EST150 151:EST151 152:EST152 153:EST153 154:EST154  
155:EST155 156:EST156 157:EST157 158:EST158 159:EST159  
160:EST160 161:EST161 162:EST162 163:EST163 164:EST164  
165:EST165 166:EST166 167:EST167 168:EST168 169:EST169  
170:EST170 171:EST171 172:EST172 173:EST173 174:EST174

175:EST175 176:EST176 177:EST177 178:EST178 179:EST179  
180:EST180 181:EST181 182:EST182 183:EST183 184:EST184  
185:EST185 186:EST186 187:EST187 188:EST188 189:EST189  
190:EST190 191:EST191 192:EST192 193:EST193  
EST-STS-THREE  
194:STS1 195:STS2 196:STS3 197:STS4 198:STS5 199:STS6  
200:STS7 201:STS8 202:STS9 203:STS10 204:STS11 205:STS12  
206:STS13 207:STS14 208:STS15 209:STS16 210:STS17  
211:STS18 212:STS19 213:STS20 214:STS21 215:STS22  
216:STS23 217:STS24 218:STS25 219:STS26 220:STS27  
221:STS28 222:STS29 223:STS30 224:STS31 225:STS32  
226:STS33 227:STS34 228:STS35 229:STS36 230:STS37  
231:STS38 232:STS39 233:STS40 234:STS41 235:STS42  
236:STS43 237:STS44 238:STS45 239:STS46 240:STS47  
241:STS48 242:STS49 243:STS50 244:STS51 245:STS52  
246:STS53 247:STS54 248:STS55 249:STS56 250:STS57

Statistics: Mean 12.474; Variance 4.448; scale 2.805

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description	Pred. No.
C 1	76	3.1	498	AA022576	ze72e10.r1 Soares fet	2.79e-46
C 2	71	2.9	829	W22594	70B5 Human retina CDN	3.45e-41
C 3	71	2.9	829	W22594	70B5 Human retina CDN	3.45e-41
C 4	61	2.5	324	W95996	yt83c02.r1 Homo sapie	3.10e-31
C 5	60	2.5	386	T09185	EST07078 Homo sapiens	2.94e-30
C 6	55	2.3	572	240	mi26g03.r1 Soares mou	1.92e-25
C 7	55	2.3	572	3	mi26g03.r1 Soares mou	1.92e-25
C 8	55	2.3	604	185	mf81d10.r1 Soares mou	1.92e-25
C 9	54	2.2	424	185	mb82e05.r1 Soares mou	1.71e-24
C 10	54	2.2	424	217	mb82e05.r1 Soares mou	1.71e-24
C 11	53	2.2	513	26	ze72e10.s1 Soares fet	1.51e-23
C 12	51	2.1	439	172	md49b07.r1 Soares mou	1.12e-21
C 13	49	2.0	372	138	S.scrofa mRNA; expires	7.98e-20
C 14	49	2.0	372	249	S.scrofa mRNA; expires	7.98e-20
C 15	45	1.9	309	180	me93d11.r1 Soares mou	3.39e-16
C 16	47	1.9	488	97	z881a08.s1 Homo sapie	5.36e-18
C 17	47	1.9	488	211	z881a08.s1 Homo sapie	5.36e-18
C 18	44	1.8	303	13	mi75d03.r1 Soares mou	2.63e-15
C 19	44	1.8	303	242	mi75d03.r1 Soares mou	2.63e-15
C 20	44	1.8	323	185	mb68b03.r1 Soares mou	2.63e-15
C 21	44	1.8	323	218	mb68b03.r1 Soares mou	2.63e-15
C 22	43	1.8	448	220	z881a08.r1 Soares sen	2.00e-14
C 23	43	1.8	448	186	z881a08.r1 Soares sen	2.00e-14
C 24	42	1.7	154	240	mi26g08.r1 Soares mou	1.50e-13
C 25	42	1.7	154	3	mi26g08.r1 Soares mou	1.50e-13
C 26	40	1.7	172	178	me65f05.r1 Soares mou	7.97e-12
C 27	36	1.5	78	185	z881a08.s1 Homo sapie	1.73e-08
C 28	36	1.5	255	36	y152d08.r1 Homo sapie	1.73e-08
C 29	37	1.5	260	43	Human aorta CDNA 5'e	2.63e-09
C 30	37	1.5	260	237	Human aorta CDNA 5'e	2.63e-09
C 31	37	1.5	273	179	me89g06.r1 Soares mou	2.63e-09
C 32	37	1.5	274	40	YF58g02.r1 Homo sapie	2.63e-09
C 33	36	1.5	306	96	YF60f09.r1 Homo sapie	1.73e-08
C 34	36	1.5	327	154	mb81f03.r1 Soares mou	1.73e-08
C 35	37	1.5	332	127	Rice CDNA, partial se	2.63e-09
C 36	37	1.5	335	16	mm32c09.r1 Stratagene	2.63e-09
C 37	37	1.5	335	244	mm32c09.r1 Stratagene	2.63e-09
C 38	36	1.5	337	196	Human STS SHGC-5727 c	1.73e-08
C 39	36	1.5	392	42	Human fetal brain CDN	1.73e-08
C 40	36	1.5	392	237	Human fetal brain CDN	1.73e-08
C 41	36	1.5	400	168	mc20f11.r1 Life Tech	1.73e-08
C 42	36	1.5	428	43	Human fetal brain CDN	1.73e-08
C 43	37	1.5	430	237	Human fetal brain CDN	2.63e-09

	Query Match	2.9%	Score 71;	DB 219;	Length 829;
	Best Local Similarity	70.4%;	Pred. No. 3.45e-41;		
	Matches 119;	Conservative 0;	Mismatches 50;	Indels 0;	Gaps 0;
Dbb	116	acttaagctcttcgattctctaaacatactcgacacctcgactcgggcangttgattttca	175		
Cp	540	ACTTGCTCTTCGGTTCTGGACCAACCTGCACCTCGGCTTCTGTGAGGTTTATTTC	481		
Dbb	176	gtccaccctctctcgcatgaagatctctgggtaccgggtctctggcaaacagtccttcca	235		
Cp	480	TGGCTAGCTCTTCTCTGGTGAAGACATCTGGGTAGTGTGTTGGGGCAAGACATGCCTCCA	421		
Dbb	236	gcacatctagctgcgcccagtgaaacgtctctctcccagngctgttt	284		

CP 420 GAGCTCCAACTGCTGAAGAGCGAAGGTTGTCGATTTCTGCGCTGTTT 372

RESULT 3 W22594 829 bp mRNA EST 06-MAY-1996  
LOCUS 7085 Human retina cDNA Tsp5091-cleaved sublibrary Homo sapiens cDNA  
DEFINITION not directional.

ACCESSION W22594  
NID 91399427

KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens

Eukaryotes; mitochondria; eukaryotes; Metazoa; Chordata; Vertebrata; Euthera; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 829)

AUTHORS Macke, J., Smallwood, P. and Nathans, J.

TITLE Adult Human Retina cDNA

JOURNAL Unpublished (1996)

COMMENT

Contact: Dr. Jeremy Nathans

Dr. Jeremy Nathans, Dept. of Molecular Biology and Genetics

Johns Hopkins School of Medicine

725 North Wolfe Street, Baltimore, MD 21205

Tel: 410 955 4678

Fax: 410 614 0827

Email: jeremy.nathans@gmail.bs.jhu.edu

Clones from this library are NOT available.

PCR Primers

FORWARD: CTTTGGACAGTTTCAGCTGTTAACT

BACKWARD: GAGTGGCTTATGAGTATTTCTTCAGGGTAA

Seq primer: GGGTAAAGCAAGAAAT.

Location/Qualifiers

1..829

/organism="Homo sapiens"

/note="Organ: eye; Vector: lambda gt10; Site\_1: EcoRI; Site\_2: EcoRI; The library used for sequencing was a sublibrary derived from a human retina cDNA library. Inserts from retina cDNA library DNA were isolated, cleaved with Tsp5091, size selected, and cloned into lambda gt10. Individual plaques were arrayed and used as templates for PCR amplification and these PCR products were used for sequencing."

/clone\_lib="Human retina cDNA Tsp5091-cleaved sublibrary"

/sex="mixed (males and females)"

/tissue\_type="retina"

/dev\_stage="adult"

/lab\_host="E. coli strain K802"

<1..829

mRNA 155 a 184 c 218 g 194 t 78 others

BASE COUNT

ORIGIN

Query Match 2.9%; Score 71; DB 186; Length 829;

Best Local Similarity 70.4%; Pred. No. 3.45e-41;

Matches 119; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

Db 116 actgctcttgcattcttaaacatcactgacctgactcggtgctgatttca 175

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

CP 540 ACTTGCTCTTCGTTCTGGAACCAACCTGCACCTGCTTCTGTGAGTTATTTC 481

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 176 gtccacacctctctgcagatgaagatgtctggtaacgggtcttggcaacagtgctcca 235

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

CP 480 TGGCTAGTCTTCTCTGTTGGAAGACATCTGGTAGTGTGTTGGGCAAGACTGCTCCA 421

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 236 gacatctagctgcgcccgagtgacgtgtctctctccggngctgttt 284

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

CP 420 GAGCTCCAACTGCTGAAGAGCGAAGGTTGTCGATTTCTGCGCTGTTT 372

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 4 R95996 324 bp mRNA EST 11-SEP-1995

LOCUS

DEFINITION Yt83c02.r1 Homo sapiens cDNA clone 230882 5' similar to gb:M93650

PAIRED BOX PROTEIN PAX-6 (HUMAN);

1.

ACCESSION

NID

KEYWORDS

SOURCE

clone=230882 primer=M13Rpl1 library=Soares pineal gland N3HPG vector=PT73D (Pharmacia) with a modified polylinker host=DH10B (ampicillin resistant) Rsite1=Not I Rsite2=Eco RI 1st strand CDNA was primed with a Not I - Oligo(dT) primer [5', TTTTACCAATCTGAAGTGGAGCGCGCGGTTTTTTTTTTTTTTT 3'], double-stranded cDNA was size selected, ligated to Eco RI adapters (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library constructed by Bento Soares and M.Fatima Bonaldo.

ORGANISM

Homo sapiens  
Eukaryotes; Metazoa; Eumetazoa; Bilateria; Coelomata; Deuterostomia; Chordata; Vertebrata; Gnathostomata; Osteichthyes; Sarcopterygii; Chonata; Tetrapoda; Amniota; Mammalia; Theria; Euthera; Archonta; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

1 (bases 1 to 324)

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M.,

Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M.,

Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F.,

Trevaskis, E., Waterston, R., Williamson, A., Wohlmann, P. and

Wilson, R.

The WashU-Merck EST Project

Unpublished (1995)

Contact: Willson RK

WashU-Merck EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: estw@wustl.edu

High quality sequence stops: 180

Source: IMAGE Consortium, LNL

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Location/Qualifiers

1..324

/organism="Homo sapiens"

/clone="230882"

<1..>324

mRNA 104 a 74 c 71 g 61 t 14 others

BASE COUNT

ORIGIN

Query Match 2.5%; Score 61; DB 125; Length 324;

Best Local Similarity 66.3%; Pred. No. 3.10e-31;

Matches 118; Conservative 0; Mismatches 60; Indels 0; Gaps 0;

Db 22 gaagcnaagcaagaataagacatcttaccacgaagcaaatgagccctggaga 81

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 367 GAAGAAACAGCGCAGAAATCGGACAACTTCGCTCTTCAGCAGTTCGAGG 426

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 82 aagagtttgagagaccattatccagatgtgttgctgagagaagactagcagccaaa 141

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 427 CAGCTTGTCCCAACACACTACCTACCCAGATGTCTTCCAGAGAAGAGTACCCATGAAA 486

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Db 142 tagatctacctgaagcaagaatacagatgtgtttcttaacgagggcgaatggag 199

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 487 TAAACCTCAGCAAGACCGAGTGCAGGTGGTTGTTCCAGACCGAGGAGGAG 544

||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5

LOCUS

DEFINITION

EST07078 Homo sapiens cDNA clone HIBBQ26 5' end similar to

Oculorhombin (aniridia).

ACCESSION T09185

NID 9390213

KEYWORDS EST.

SOURCE

Human clone-HIBBQ26 library-Infant brain, Bento Soares primer-W13

Reverse.

ORGANISM

Homo sapiens

Eucaryotes; Metazoa; Chordata; Vertebrata; Gnathostomata; Mammalia;



```

Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 460.
Location/Qualifiers
1..604
/organism="Mus musculus"
/strain="C57BL/6J"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5'
TGTTACCAATCTGAAGTCGGAGCGCCGGAATTTTTTTTTTTTTTTTTT
T 3'] , on equal amounts of mRNA from 2 13.5dpc and 2
14.5dpc embryos [total RNA provided by Minoru Ko, Wayne
State Univ., from 2 ]; double-stranded cDNA was ligated to
Eco RI adaptors (Pharmacia), digested with Not I and
cloned into the Not I and Eco RI sites of the modified
pT73 vector. Library went through one round of
normalization, and was constructed by Bento Soares and
M.Fatima Bonaldo."
/cclone="420691"
/clone_lib="Soares mouse embryo NBMEL3.5 14.5"
/sex="unknown"
/tissue_type="embryo"
/dev_stage="13.5-14.5dpc total fetus"
/lab_host="DH10B"
<!-->604
130 a 200 c 171 g 102 t 1 others
mrna
BASE COUNT
ORIGIN
Query Match      2.3%; Score 55; DB 182; Length 604;
Best Local Similarity 66.3%; Pred. No. 1.92e-25;
Matches 112; Conservative 0; Mismatches 57; Indels 0; Gaps 0;

Db    113 aagcagcgccggaatgaaccattcaacagcagcgccttcaggcgctggagcggtga 172
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY   372 AAACAGCGCAGAAATCGGCACACTTCGTCTTCACAGTTGGAAAGCTCTGGAGGAGTC 431

Db    173 tttagcgccacacactcacctcagcgctttgtgctaagagctagtgcgcgtgtcaac 232
      ||||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY   432 TTTGCCAACAACACTACCAGATGTCTTCACGAGAGAAGAGCTAGCCATAAATAAAC 491

Db    233 ctcaatgaagcagctgtccaagtctggttccagaaccgcgcgtgccaaagt 281
      |||| | | | | | | | | | | | | | | | | | | | | | | | | | |
QY   492 CTCACAGAAGCCAGAGTGCAGGTTTGTTCCAGAACCCGAAGCCCAAGT 540

RESULT          9
LOCUS           W17990             424 bp        mRNA            EST            10-SEP-1996
DEFINITION      mb82e05.r1 Soares mouse p3NMf19.5 Mus musculus cDNA clone 335936 5'
                similar to gb:U03873 Mus musculus homeobox (MOUSE);.
ACCESSION       W17990
NID             Q1292374
KEYWORDS         EST.
SOURCE           house mouse.
ORGANISM        Mus musculus
                Eukaryotes; mitochondrial eukaryotes; Metazoa; Chordata;
                Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
                Mus.
REFERENCE       1 (bases 1 to 424)
AUTHORS         Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
                Geisels,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
                Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
                Theising,B., Wyllie,T., Lennon,G., Soares,B., Wilson,R. and
                Waterston,R.
TITLE           The WashU-HMMI Mouse EST Project
JOURNAL          Unpublished (1996)
COMMENT         Contact: Marra M/Mouse EST Project
                 WashU-HMMI Mouse EST Project
                 Washington University School of MedicineP
                 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
                 Tel: 314 286 1800
                 Fax: 314 286 1810
                 Email: mouseest@watson.wustl.edu

```

```

FEATURES
  Location/Qualifiers
    1..513
    /organism="Homo sapiens"
    high quality sequence stop: 448.

```



MGI:233133  
Seq primer: ETPRimer  
High quality sequence stop: 347.  
Location/Qualifiers

FEATURES

source

1..439  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTCACCATCTGAAGTGCGAGCGCGGCGAAATTTTTTTTTTTTTTTT T 3'], on equal amounts of mRNA from 2 13.5dpc and 2 14.5dpc embryos [total RNA provided by Minoru Ko, Wayne State Univ., from 2 ]; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."  
/clone="371701"  
/clone\_lib="Soares mouse embryo NDMEI3.5 14.5"  
/sex="unknown"  
/tissue.type="embryo"  
/dev\_stage="13.5-14.5dpc total fetus"  
/lab\_host="DH10B"  
<1..>439

mRNA  
BASE COUNT 119 a 124 c 132 g 64 t  
ORIGIN

Query Match 2.18; Score 51; DB 172; Length 439;  
Best Local Similarity 66.74; Pred. N.1.12e-21;  
Matches 102; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

Ddb 170 ttactagccagcagctcgaggagtgcggtagcaccatttcagagaaacgctaccagac 229  
||| || ||||| ||| ||||| ||| || | ||||| |||  
Q2Y 396 TTCCTCTTCAGCAGTGGNAGCTCTGGAGGCAGCTCTTGCCCAACACATACCAGAT 455  
||| || ||||| ||| ||||| ||| || | ||||| |||  
Ddb 230 atgtccactcgcgaagaataacgcgtgtgaccaccttacggaagcccgcggttt 289  
||| || ||||| ||| ||||| ||| || | ||||| |||  
Q2Y 456 GTCTTCACGAGAAGAGCTAGCCATGAATAAACCCTCACAGAAGCCAGATGCCAGTT 515  
||| || ||||| ||| ||||| ||| || | ||||| |||  
Ddb 290 tggttcaaatcgccgggccaaatggagaaaag 322  
||| || ||||| ||| ||||| ||| || | ||||| |||  
Q2Y 516 TGGTTCAGAACCCGAGAGCCAAAGTGAGGGAAG 548  
||| || ||||| ||| ||||| ||| || | ||||| |||

RESULT 13  
LOCUS SSC1E11 372 bp RNA EST 03-SEP-1996  
DEFINITION S.scrofa mRNA; expressed sequence tag (5'; clone clell).  
ACCESSION F14521  
KEYWORDS g971739  
SOURCE EST; expressed sequence tag.  
ORGANISM pig.  
Sus scrofa  
Eukaryote; mitochondrial eukaryotes; Metazoa; Chordata;  
Vertebrata; Eutheria; Artiodactyla; Suiformes; Sulina; Suidae; Sus.  
REFERENCE 1 (bases 1 to 372)  
AUTHORS Winteroe,A.K., Fredholm,M. and Davies,W.  
TITLE Evaluation and characterization of a porcine small intestine cDNA library  
JOURNAL Mamm. Genome 7, 509-517 (1996)  
REFERENCE 2 (bases 1 to 372)  
AUTHORS Winteroe,A.K.  
TITLE Direct Submission  
JOURNAL Submitted (26-JUL-1995) Winteroe A.K., The Royal Veterinary and Agricultural University, Department of Animal Science and Animal Health, Division of Animal Genetics, Bulowsvej 13, 1870 Frederiksberg C, DENMARK  
LOCATION/Qualifiers 1..372  
/organism="Sus scrofa"  
/tissue.type="small intestine"  
/clone\_lib="directionally cloned cDNA in XLI-blue MRF."

FEATURES

source

QY	436	CCCAACACACTCCAGAGTGTCTTCCACGAGAGAGCTAGCCATGAATAAACCTCA	495
Db	320	cagaagctcgggtacagatctgttcagantcagcagcagcagcgcgag 372	
QY	496	CAGAGCCAGAGTGCAGGTTGGTTCAGAACCCAGAGAGCAAGTGGAGGAAG	548
RESULT	15		
LOCUS		309 bp mRNA EST	12-SEP-1996
DEFINITION		me93d11.r1 Soares mouse embryo NbME13.5 14.5 Mus musculus CDNA	
		(cloned 403125 5' similar to gb:52875 Mouse homeobox gene S8 mRNA	
ACCESSION		(MOUSE);	
NID		W81935	
KEYWORDS		g1392750	
SOURCE		EST.	
ORGANISM		house mouse.	
		Mus musculus	
		Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata;	
		Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;	
		Mus.	
REFERENCE		1 (bases 1 to 309)	
AUTHORS		Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,	
		Geisler,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,	
		Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,	
		Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and	
		Waterston,R.	
TITLE		The WashU-HMI Mouse EST Project	
JOURNAL		Unpublished (1996)	
COMMENT			
		Contact: Marra M/Mouse EST Project	
		WashU-HMI Mouse EST Project	
		Washington University School of Medicine	
		4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108	
		Tel: 314 286 1800	
		Fax: 314 286 1810	
		Email: mouseest@wustl.edu	
		This clone is available royalty-free through LLNL; contact the	
		IMAGE Consortium (info@image.llnl.gov) for further information.	
		MGI:246893	
		Trace considered overall poor quality	
		Seq primer: -28M13 rev2 from Amersham	
		High quality sequence stop: 1.	
FEATURES			
Source		Location/Qualifiers	
		1..309	
		/organism="Mus musculus"	
		/strain="C57BL/6J"	
		/note="Vector: pT7T3D-Pac (Pharmacia) with a modified	
		polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA	
		was primed with a Not I - oligo(dT) primer [5,	
		TGTTACCATCTGAGTGGCGCGCGGAAATTTTTTTTTTTTTTTTTTTT	
		T 3'], on equal amounts of mRNA from 2 13.5dpc and 2	
		14.5dpc embryos [total RNA provided by Minoru Ko, Wayne	
		State Univ., from 2 ]; double-stranded cDNA was ligated to	
		Eco RI adaptors (Pharmacia) digested with Not I and	
		cloned into the Not I and Eco RI sites of the modified	
		pT7T3 vector. Library went through one round of	
		normalization, and was constructed by Bento Soares and	
		M.Fatima Bonaldo."	
		/clone="403125"	
		/clone_lib="Soares mouse embryo NbME13.5 14.5"	
		/sex="unknown"	
		/tissue_type="embryo"	
		/dev_stage="13.5-14.5dpc total fetus"	
		/lab_host="DH10B"	
		<1..>309	
BASE COUNT		55 a 84 c 92 g 78 t	
ORIGIN			
Query Match		1.9%	Score 45; DB 180; Length 309;
Best Local Similarity		64.7%	Pred. No. 3.39e-16;
Matches 110; Conservative		0; Mismatches 59; Indels 1; Gaps 1;	

Db 99 actttgacggcggttcttggaaccagacttggaacacgtgcttcaactgatgttgacacggc 158  
||| | | ||||| ||||| || | | ||||| ||| |||  
Cp 540 ACTTGGCTCTTCGGTTCGTGAACCAACCACTGCACTCTGGCTTCTGTGAGTTTATTTC 481  
Db 159 gagctagctcttcacgcacaaggcgtcagggttagttagtgcgcctcaatacacgctcc 218  
||| ||||| ||||| | | | | | ||||| ||| |||  
Cp 480 TGGCTAGCTCTCTCTGTGTGAAGACATCTGGGT-AGTGTGTTTGGGCAAGACTGCCTCC 422  
Db 219 tggccttgacgtggtgctgttgaatggttgattccggcgctgctt 268  
| ||| | | ||| ||| ||| ||| ||| ||| |||  
Cp 421 AGAGCTTCCAACTGCTGAAGAGCGAAGCTTGTCCGATTTCTGGCGTGTTT 372

Search completed: Fri May 30 16:40:55 1997  
Job time : 2071 secs.

dla.

\*\*\*\*\*  
WATERMAN  
\*\*\*\*\*  
(TM)  
\*\*\*\*\*

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (c) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by Intelligent, Inc.

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm  
Run on: Thu May 29 11:34:49 1997; MasPar time 13.15 Seconds  
abular output not generated. 570.251 Million cell updates/sec

Title: >US-08-701-278-2  
Description: (1-263) from US08701278.pep  
Perfect Score: 1857  
Sequence: 1 MFVHCPPQLEGTAPGNHS.....EGSQDKPSPTKEQSEGEKSV 263

Scoring table: PAM 150  
Gap 11  
Searched: 89912 seqs, 28507787 residues  
Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: p1r50  
1:ann1 2:ann2 3:ann3 4:ann4 5:unann1 6:unann2 7:unann3  
8:unann4 9:unann5 10:unann6 11:unann7 12:unann8  
13:unann9 14:unann10 15:unenc 16:unrev  
Statistics: Mean 44.651; Variance 107.204; scale 0.417

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES				
Result No.	Score	Query Match %	ID	Description
1	399	21.5	A46403	transcription factor
2	392	21.1	I48713	Phox2 homeodomain pr
3	375	20.2	I50413	paired-related homeo
4	374	20.1	I53118	MHOX - mouse
5	374	20.1	I48902	homeobox protein - m
6	374	20.1	I52076	DNA-binding protein
7	373	20.1	I51226	homeodomain protein
8	373	20.1	I27842	homeotic protein smo
9	371	20.0	I48185	gene alx3 protein -
10	369	19.9	I45452	transcription factor
11	370	19.9	I50115	transcription factor
12	369	19.9	I51031	pax-3 protein - mous
13	366	19.7	I51031	homeotic protein S8
14	366	19.7	I48410	gene S8 protein - mo
15	366	19.7	I53440	Prx-2 protein - chic
16	364	19.6	I60250	mab-18 protein (tran
17	364	19.6	I60251	mab-18 protein (tran
18	364	19.6	I60252	vab-3 protein - Caen
19	363	19.5	I47523	cartilage homeoprote
20	360	19.4	I49265	Pax7 - mouse (fragme
21	354	19.1	I51234	gene Pax-6 protein -

22	354	19.1	269 14	S36166	Pax-6 protein - rat	6.90e-37
23	354	19.1	416 13	S37689	Pax-QNR protein - qu	6.90e-37
24	354	19.1	422 13	A56674	paired box homeotic	6.90e-37
25	354	19.1	436 14	S42234	pax-6 protein - mous	6.90e-37
26	354	19.1	447 13	A41644	homeotic protein ani	6.90e-37
27	353	19.0	437 12	S18558	pax protein - zebra	9.97e-37
28	353	19.0	442 12	A57374	homeotic protein PAX	9.97e-37
29	353	19.0	613 12	A26062	segmentation protein	9.97e-37
30	351	18.9	361 14	I49594	homeobox protein - m	2.08e-36
31	348	18.7	80 12	A26332	homeotic protein BSH	6.23e-36
32	347	18.7	449 12	B43698	BSH4 protein - fruit	8.99e-36
33	346	18.6	184 12	S20258	homeotic protein ceh	1.30e-35
34	342	18.4	245 7	A47539	gastrulation-express	5.60e-35
35	342	18.4	288 12	I51620	homeobox protein - A	5.60e-35
36	341	18.4	838 12	I45557	eyeless, long form -	8.07e-35
37	339	18.3	289 14	S35346	otx2 protein - mouse	1.67e-34
38	338	18.2	252 7	A54677	homeotic protein goo	2.41e-34
39	338	18.2	256 7	A42768	gastrulation-express	2.41e-34
40	338	18.2	355 14	I56547	homeodomain 159..341	2.41e-34
41	338	18.2	427 12	A43698	BSH9 protein - fruit	2.41e-34
42	336	18.1	323 16	I50509	orthodenticle-relate	5.00e-34
43	335	18.0	243 12	I51424	homeobox protein - A	7.20e-34
44	334	18.0	243 7	B42768	homeotic protein goo	1.04e-33
45	333	17.9	612 12	A54282	reversed polarity pr	1.49e-33

ALIGNMENTS

RESULT 1  
ENTRY transcription factor with prd-type homeo domain and  
TITLE Pro/Gln-rich domain-al - Drosophila  
ORGANISM #formal\_name Drosophila  
DATE 21-Sep-1993 #sequence\_revision 21-Sep-1993 #text\_change 19-Nov-1993  
ACCESSIONS A46403  
REFERENCE A46403  
#authors Schneitz, K.; Spielmann, P.; Noll, M.  
#journal Genes Dev. (1993) 7:114-129  
#title Molecular genetics of aristaless, a prd-type homeo box gene involved in the morphogenesis of proximal and distal pattern elements in a subset of appendages in Drosophila.  
#cross-references MUID:93138380  
#accession A46403  
#status preliminary  
#molecule\_type nucleic acid  
#residues #molecules 1-384 #label SCH  
#cross-references NCBI:123442; NCBI:123444  
#note sequence extracted from NCBI backbone  
SUMMARY #length 384 #molecular-weight 41058 #checksum 2627

```
#cross-references MUID:93138380
#accession A46403
#status preliminary
#molecule_type nucleic acid
#residues 1-384 ##label SCH
#cross-references NCBIN:123442; NCBIPI:123444
#note sequence extracted from NCBI backbone
SUMMARY #length 384 #molecular-weight 41058 #checksum 2627

Query Match 21.5%; Score 399; DB:16; Length 384;
Best Local Similarity 71.8%; Pred. NO. 4.05e-44;
Matches 51; Conservative 9; Mismatches 11; Indels 0; Gaps 0;

Db 82 krkrryrttftsfqleekafsrthypdvftreelamkigtleariqvfwfqrakr 141
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
QY 30 RRKQRNRTTFALQQLALEAVFAQTHYPDVFTREELAMKINLTARVQVWFQNRRAKR 89
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

Db 142 kqekvgppqshp 152
|:|:|:|
QY 90 KTERGASDQEP 100

RESULT 2
ENTRY
TITLE #type complete
ORGANISM Phox2 homeodomain protein - mouse
DATE #formal_name Mus musculus #common_name house mouse
02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 02-Jul-1996
ACCESSIONS I48713
REFERENCE I48314
#authors Valarche, I.; Tissier-Seta, J.P.; Hirsch, M.R.; Martinez, S.;
```

3

2

```

##residues 1-28,29-306 ##label HOT
##cross-references NCBIN:126845; NCBIP:126846; NCBIN:126847;
NCBIP:126848
#note sequence modified after extraction from NCBI backbone
REFERENCE A56744
#authors Macina, R.A.; Barr, F.G.; Galili, N.; Riehlman, H.C.
#journal Genomics (1995) 26:1-8
#title Genomic organization of the human PAX3 gene: DNA sequence analysis of the region disrupted in alveolar rhabdomyosarcoma.
#accession A56744
#molecule_type DNA
#residues 1-28;43-326 ##label MAC
##cross-references GB:U12263
GENETICS
#gene GDB:PAX3; WS1
#map_position 2q36-2q36
#cross-references GDB:120495
CLASSIFICATION #superfamily homeobox homology
FEATURE
67-123 #domain homeobox homology #label HOX
#length 326 #checksum 8075
#summary
Query Match 19.9%; Score 369; DB 13; Length 326;
Best Local Similarity 66.7%; Pred. No. 2.76e-39;
Matches 46; Conservative 13; Mismatches 9; Indels 1; Gaps 1;
Db 62 lkrqrrrtftaeqlerakferthpdytreelaqrakltearvqvwnrrarw 121
QY 29 LRRKORRRRTTFAQQLEALEAVFAQTHYPDVFTRTELAMKINLTARVQVWFQNRRAKW 88
Db 122 rk-qagang 129
QY 89 RTERGASD 97

RESULT 11
ENTRY S50115 #type fragment
TITLE transcription factor pax-7 - human (fragment)
ORGANISM #formal_name Homo sapiens #common_name man
DATE 14-Jul-1995 #sequence_revision 10-Nov-1995 #text_change
19-Jan-1996
ACCESSIONS S50115
REFERENCE Schaefer, B.W.; Czerny, T.; Bernasconi, M.; Genini, M.; Busslinger, M.
#journal Nucleic Acids Res. (1994) 22:4574-4582
#title Molecular cloning and characterization of a human PAX-7 cDNA expressed in normal and neoplastic myocytes.
#accession S50115
#status preliminary
#molecule_type DNA
#residues 1-467 ##label SCH
#cross-references EMBL:235141
CLASSIFICATION #superfamily paired box homology; homeobox homology
FEATURE
34-161 #domain paired box homology #label PBH\
218-274 #domain homeobox homology #label HOX
#length 467 #checksum 4440
#summary
Query Match 19.9%; Score 370; DB 13; Length 467;
Best Local Similarity 66.7%; Pred. No. 1.91e-39;
Matches 46; Conservative 13; Mismatches 9; Indels 1; Gaps 1;
Db 213 lkrqrrrtftaeqlerakferthpdytreelaqrakltearvqvwnrrarw 272
QY 29 LRRKORRRRTTFAQQLEALEAVFAQTHYPDVFTRTELAMKINLTARVQVWFQNRRAKW_88
Db 273 rk-qagang 280
QY 89 RTERGASD 97

```

```

RESULT 12
ENTRY S15031 #type complete
TITLE pax-3 protein - mouse
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 21-Nov-1993 #sequence_revision 21-Jul-1995 #text_change
08-Dec-1995
ACCESSIONS S15031
REFERENCE S15031
#authors Goulding, M.D.; Chalepakis, G.; Deutsch, U.; Erselius, J.R.; Gruss, P.
#journal EMBO J. (1991) 10:1135-1147
#title Pax-3, a novel murine DNA binding protein expressed during early neurogenesis.
#cross-references MUID:91216108
#accession S15031
#status preliminary
#molecule_type mRNA
#residues 1-479 ##label GOU
CLASSIFICATION #superfamily paired box homology; homeobox homology
FEATURE
34-159 #domain paired box homology #label PBH\
220-276 #domain homeobox homology #label HOX
#length 479 #molecular_weight 52984 #checksum 6402
#summary
Query Match 19.9%; Score 369; DB 14; Length 479;
Best Local Similarity 66.7%; Pred. No. 2.76e-39;
Matches 46; Conservative 13; Mismatches 9; Indels 1; Gaps 1;
Db 215 lkrqrrrtftaeqlerakferthpdytreelaqrakltearvqvwnrrarw 274
QY 29 LRRKORRRRTTFAQQLEALEAVFAQTHYPDVFTRTELAMKINLTARVQVWFQNRRAKW 88
Db 275 rk-qagang 282
QY 89 RTERGASD 97

RESULT 13
ENTRY S18038 #type fragment
TITLE homeotic protein S8 - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change
30-Sep-1993
ACCESSIONS S18038
REFERENCE S18038
#authors Opstelten, D.J.E.; Vogels, R.; Robert, B.; Kalkhoven, E.; Zwartkruis, F.; de Laaf, L.; Destree, O.H.; Deschamps, J.; Lawson, K.A.; Meijlink, F.
#submission submitted to the EMBL Data Library, May 1990
#description The mouse homeobox gene, S8, is expressed during embryogenesis predominantly in mesenchyme.
#accession S18038
#molecule_type mRNA
#residues 1-164 ##label OPS
#cross-references EMBL:X52875
CLASSIFICATION #superfamily homeobox homology;
KEYWORDS DNA binding; homeobox
FEATURE
16-72 #domain homeobox homology #label HOX
#length 164 #checksum 4472
#summary
Query Match 19.7%; Score 366; DB 14; Length 164;
Best Local Similarity 70.8%; Pred. No. 8.34e-39;
Matches 46; Conservative 9; Mismatches 10; Indels 0; Gaps 0;
Db 12 kkkqrtrrtftaeqlerakferthpdytreelaqrakltearvqvwnrrarw 71
QY 30 RRRKORRRRTTFAQQLEALEAVFAQTHYPDVFTRTELAMKINLTARVQVWFQNRRAKW 89
Db 72 rnera 76
QY 90 KTERG 94

```



Search completed: Thu May 29 11:35:54 1997  
Job time : 65 secs.

```
RESULT 14
ENTRY I48410 #type fragment
TITLE gene S8 protein - mouse (fragment)
ORGANISM #formal_name Mus musculus #common_name house mouse
DATE 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change
02-Jul-1996
ACCESSIONS I48410
REFERENCE I48410
#authors Opstelten, D.J.; Vogels, R.; Robert, B.; Kalkhoven, E.;
Zwartkruis, F.; de Laaf, L.; Destree, O.H.; Deschamps, J.;
Lawson, K.A.; Weijlink, F.
#journal Mech. Dev. (1991) 34:29-41
#title The mouse homeobox gene, S8, is expressed during
embryogenesis predominantly in mesenchyme.
#cross-references MUID:92001534
#accession I48410
#status preliminary; translated from GB/EMBL/DBJ
#molecule_type mRNA
#residues 1-164 #label RES
#cross-references EMBL:X52875; NID:g51365; CDS_PID:g51366
JENETICS
#note gene name S8
SUMMARY #length 164 #checksum 4472

Query Match 19.7%; Score 366; DB 14; Length 164;
Best Local Similarity 70.8%; Pred. No. 8.34e-39;
Matches 46; Conservative 9; Mismatches 10; Indels 0; Gaps 0;

Db 12 kkkqrntttfssqlqalervferthypdafvrselarrvnlsearvqvfgnrrakfr 71
Qy 30 RRKQRNRTTALQOLEALEAVFAQTHYPDPVFTREELAMKINLTARVQVWFQNRRAKWR 89

Db 72 rnera 76
Qy 90 KTERG 94

RESULT 15
ENTRY S49440 #type complete
TITLE Prx-2 protein chicken
ORGANISM #formal_name Gallus gallus #common_name chicken
DATE 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change
07-Jul-1995
ACCESSIONS S49440
REFERENCE S49440
#authors Brouwer, A.; Weijlink, F.
#submission Submitted to the EMBL Data Library, June 1994
#description Expression of members of the Prx-family of homeobox genes (S8
and Mhox/K2/Pmx/Phox1) in mouse and chicken embryos.
#accession S49440
#status preliminary
#molecule_type mRNA
#residues 1-165 #label BRO
#cross-references EMBL:X79695
CLASSIFICATION #superfamily homeobox homology
FEATURE
17-73 #domain homeobox homology #label HOX
SUMMARY #length 165 #molecular-weight 18691 #checksum 7828

Query Match 19.7%; Score 366; DB 13; Length 165;
Best Local Similarity 70.8%; Pred. No. 8.34e-39;
Matches 46; Conservative 9; Mismatches 10; Indels 0; Gaps 0;

Db 13 kkkqrntttfssqlqalervferthypdafvrselarrvnlsearvqvfgnrrakfr 72
Qy 30 RRKQRNRTTALQOLEALEAVFAQTHYPDPVFTREELAMKINLTARVQVWFQNRRAKWR 89

Db 73 rnera 77
Qy 90 KTERG 94
```

\*\*\*\*\*  
W A R E N  
\*\*\*\*\* (TM)

Release 2.1D John F. Collins, Biocomputing Research Unit.  
Copyright (C) 1993, 1994, 1995 University of Edinburgh, U.K.  
Distribution rights by Intelligenetics, Inc.

MPsrch\_pp protein - protein database search, using Smith-Waterman algorithm

Run on: Thu May 29 11:33:55 1997; MasPar time 9.24 Seconds  
603.427 Million cell updates/sec

abular output not generated.

Title: >US-08-701-278-2  
Description: (1-263) from US08701278.pep  
Perfect Score: 1857  
Sequence: 1 MFYFHCPPQLGTAPEGNHS.....EGSQDKPSPPTKEQGEKSV 263

Scoring table: PAM 150  
Gap 11

Searched: 59021 seqs, 21210388 residues

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database: swiss-prot34  
1:part1 2:part2 3:part3 4:part4 5:part5 6:part6 7:part7  
8:part8 9:part9 10:part10 11:part11

Statistics: Mean 46.324; Variance 87.195; scale 0.531

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

sult No.	Score	Match	Length	ID	Description	Pred. No.
1	375	20.2	244	5	HPRI_CHICK HOMEBOX PROTEIN PRX-	2.18e-51
2	374	20.1	199	7	PMX1_HUMAN PAIRED MESODERM HOMO	3.49e-51
3	374	20.1	217	7	PMX1_MOUSE PAIRED MESODERM HOMO	3.49e-51
4	370	19.9	467	7	PAX7_HUMAN PAIRED BOX PROTEIN PA	2.27e-50
5	369	19.9	479	7	PAX3_MOUSE PAIRED BOX PROTEIN PA	3.62e-50
6	369	19.9	479	7	PAX3_HUMAN PAIRED BOX PROTEIN PA	3.62e-50
7	360	19.4	290	7	PAX7_MOUSE PAIRED BOX PROTEIN PA	2.42e-48
8	357	19.2	419	4	GSC_DRONE HOMEBOX PROTEIN GOOS	9.79e-48
9	354	19.1	216	7	PAX6_CHICK PAIRED BOX PROTEIN PA	3.95e-47
10	354	19.1	416	7	PAX6_COTJA PAIRED BOX PROTEIN PA	3.95e-47
11	354	19.1	422	7	PAX6_MOUSE PAIRED BOX PROTEIN PA	3.95e-47
12	354	19.1	422	7	PAX6_HUMAN PAIRED BOX PROTEIN PA	3.95e-47
13	353	19.0	437	7	PAX6_BRARE PAIRED BOX PROTEIN PA	6.29e-47
14	353	19.0	613	5	HMPL_DRONE SEGMENTATION PROTEIN	6.29e-47
15	351	18.9	449	4	GSBP_DRONE GOOSEBERRY PROXIMAL P	1.59e-46
16	346	18.6	252	10	UNX4_CAEEL HOMEBOX PROTEIN UNC-	1.62e-45
17	342	18.4	245	4	GSC_CHICK HOMEBOX PROTEIN GOOS	1.03e-44
18	339	18.3	289	7	OTX2_MOUSE HOMEBOX PROTEIN OTX2	4.13e-44
19	338	18.2	256	4	GSC_MOUSE HOMEBOX PROTEIN GOOS	6.56e-44
20	338	18.2	289	7	OTX2_HUMAN HOMEBOX PROTEIN OTX2	6.56e-44
21	338	18.2	427	4	GSBD_DRONE GOOSEBERRY DISTAL PRO	6.56e-44
22	337	18.1	240	4	GSC_BRARE HOMEBOX PROTEIN GOOS	1.04e-43

ID	HPRI_CHICK	STANDARD;	PRT;	244 AA.
AC	Q05437;			
DT	01-JUN-1994 (REL. 29, CREATED)			
DT	01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)			
DT	01-OCT-1994 (REL. 30, LAST ANNOTATION UPDATE)			
DE	HOMEBOX PROTEIN PRX-1.			
GN	PRX-1			
OS	GALLUS GALLUS (CHICKEN).			
OC	EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;			
OC	GALLIFORMES.			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE-LIMB BUD.			
RX	MEDLINE; 93321789.			
RA	NOHNO T., KOTAMA E., MYOKAI F., TANIGUCHI S., OHUCHI H., SAITO T.,			
RA	NOJI S.;			
RL	DEV. BIOL. 158:254-264(1993).			
CC	-!- FUNCTION: MAY BE INVOLVED IN THE DIFFERENTIATION OF BONE, MUSCLE,			
CC	AND OTHER TISSUES OF MESODERMAL ORIGIN DURING LIMB DEVELOPMENT.			
CC	-!- SUBCELLULAR LOCATION: NUCLEAR.			
CC	-!- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN THE MESODERMAL			
CC	CELLS OF THE LIMB BUD, VISCERAL ARCHES AND CRANIOFACIAL PROCESS,			
CC	AND AT LOWER LEVELS, IN CRANIAL MESENCHYME, UPPER AND LOWER			
CC	EYELEIDS, SOMITES AND CARTILAGE OF VERTEBRA.			
CC	-!- SIMILARITY: STRONG WITH OTHER "PAIRED-TYPE" HOMEBOX PROTEINS.			
DR	EMBL; D13433; G222851; -			
DR	HSSP; P02836; 1ENH.			
DR	PROSITE; PS00027; HOMEBOX.			
KW	HOMEBOX; DNA-BINDING; DEVELOPMENTAL PROTEIN; NUCLEAR PROTEIN.			
FT	DNA BIND 94 153			
SQ	SEQUENCE 244 AA; 27130 MW; 6BD4C3A2 CRC32;			
Query Match 20.2%; Score 375; DB 5; Length 244;				
Best Local Similarity 55.3%; Pred. No. 2.18e-51;				
Matches 52; Conservative 16; Mismatches 25; Indels 1; Gaps 1;				
Db	91 krkqrntttfnsqqlarvferthypdafvredlarrvnltearvqvfnrakfr 150			
Qy	:			
Qy	30 RRKQRNRTTALQLEALEAVFAQTHYPDVTREELAMKINLTARVQVFNRAKWR 89			
Db	151 rneimlasknasllksygdvtaveqivpripap 184			
Qy	:			
Qy	90 KTERGASDQEPGA-KEPMAEVTTPPVNRINSPPP 122			

```
CC CC MESODERMAL MUSCLE TYPES. THE PROTEIN BINDS TO AN A/T-RICH ELEMENT  
CC IN THE MUSCLE CREATINE NUCLEAR.  
CC -I- SUBCELLULAR LOCATION: CYTOSOL;  
CC -I- TISSUE SPECIFICITY: EXPRESSED EXCLUSIVELY IN MESODERALLY DERIVED  
CC CELL TYPES. DURING EMBRYOGENESIS, HIGHEST LEVELS OF EXPRESSION  
CC ARE FOUND IN THE MESENCHYME AND PRECARTILAGE ELEMENTS OF THE FACE  
CC AND HAND LIMBS. IN THE ADULT, EXPRESSION IS RESTRICTED TO SKELETAL  
CC MUSCLE, HEART AND UTERUS.  
CC -I- DEVELOPMENTAL STAGE: EXPRESSED DURING CARIOGENESIS.  
CC -I- ALTERNATIVE PRODUCTS: TWO FORMS, PMX1-A (SHOWN HERE) AND PMX1-B,  
CC ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -I- SIMILARITY: STRONG WITH OTHER "PAIRED-TYPE" HOMEOBOX PROTEINS.  
CC DR EMBL; L06502; G199584; -  
CC DR EMBL; U03873; G460135; -  
CC DR EMBL; X59725; G51362; -  
CC KW PROSITE; PS00027; HOMEOBOX.  
CC DR HOMEOBOX; DNA-BINDING; DEVELOPMENTAL PROTEIN; NUCLEAR PROTEIN;  
KW ALTERNATIVE SPLICING; PHOSPHORYLATION.  
FT FT DNA_BIND 94 153  
FT MOD_RES 197 197  
FT VARSPLIC 200 217  
FT SEQUENCE 217 AA; 24369 MW; 66A77F60 CRC32;  
  
Query Match 20.1%; Score 374; DB 7; Length 217;  
Best Local Similarity 54.7%; Pred. No. 3.49e-51;  
Matches 52; Conservative 18; Mismatches 23; Indels 2; Gaps 3;  
  
Db 91 krkqrnrttfnssqlqalervferthydpdafvredlarrvnltcarqvfwgnrakfr 150  
QY :II::II::II::II::II::II::II::II::II::II::II::II::II::II::II::II::II:  
30 RRKRRNRRTTFALOQLAEAVEAFQTHYPDVFTREELAMKINTEARVQVFQNRAKWR 89  
  
Db 151 rneramlanknasllksygdvtaveqpivprpap 185  
QY :III::II::II::II::II::II::II::II::II::II::II::II::II::II::II:  
90 KTERGA-SDQEPGA-KEPMAEVTPPPVRNINSPPP 122  
  
RESULT 4  
ID PAX7_HUMAN STANDARD; PRT; 467 AA.  
AC P23759;  
DT 01-NOV-1991 (REL. 20, CREATED)  
DI 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
DT 01-NOV-1995 (REL. 32, LAST ANNOTATION UPDATE)  
DE PAIRED BOX PROTEIN PAX-7 (HUP1).  
GN PAX7  
GD OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 95075634.  
RA SCHAEFER B.W., CZERNY T., BERNASCONI M., GENINI M., BUSSLINGER M.;  
RL NUCLEIC ACIDS RES. 22:4574-4582(1994).  
[2]  
RN SEQUENCE OF 30-195 FROM N.A.  
RM MEDLINE; 89305521.  
RX BURRI M., TRONVOUKIS Y., BOPP D., FRIGERIO G., NOLL M.;  
RL EMBO J. 8:1183-1190(1989).  
CC -I- FUNCTION: PROBABLE TRANSCRIPTION FACTOR. IT MAY HAVE A ROLE IN  
CC MYOGENESIS.  
CC -I- SUBCELLULAR LOCATION: NUCLEAR.  
CC -I- SUBUNIT: CAN BIND AS A HETERODIMER WITH PAX3.  
CC -I- SIMILARITY: STRONG WITH OTHER "PAIRED-TYPE" HOMEOBOX PROTEINS.  
CC -I- SIMILARITY: CONTAINS A PAIRED BOX DOMAIN.  
DR EMBL; Z35141; G602343; -  
DR EMBL; X15042; E14747; -  
DR EMBL; X15250; E14749; -  
DR EMBL; X15251; E14751; -  
DR PIR; S06959; S06959.  
DR MIM; 167410; -  
DR PROSITE; PS00027; HOMEOBOX.  
DR PROSITE; PS00034; PAIRED_BOX.
```

DR TRANSFAC; T00396; --  
 KW TRANSCRIPTION REGULATION; HOMEOBOX; DNA-BINDING; NUCLEAR PROTEIN;  
 KW DEVELOPMENTAL PROTEIN; PAIRED BOX.  
 FT DOMAIN 34 161 PAIRED BOX.  
 FT DNA\_BIND 217 276 HOMEOBOX.  
 FT DOMAIN 340 346 POLY-ALA.  
 FT CONFLICT 151 152 MISSING (IN REF. 2).  
 SQ SEQUENCE 467 AA; 51019 MW; D2C7BCBD CRC32;

Query Match 19.9%; Score 370; DB 7; Length 467;  
 Best Local Similarity 66.7%; Pred. No. 2,27e-50;  
 Matches 46; Conservative 13; Mismatches 9; Indels 1; Gaps 1;

Db 213 lkrqrrttftaelelekaferthypdiyltreelaqrktlrearvqvwfsnriarw 272  
 :|||||:|||||: ||| | | :|||||:|||||: ||||||| |||||:  
 QY 29 LRRKQRRNRTTFALQQLALEAVFAQTHYDPVFTREELAMKINLTREARVQVWFQNRRAK 88

Db 273 rk-qagang 280

QY 89 RKTERGASD 97

RESULT 5  
 ID PAX3\_MOUSE STANDARD; PRT; 479 AA.  
 AC P24610;  
 DT 01-MAR-1992 (REL. 21, CREATED)  
 DT 01-MAR-1992 (REL. 21, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE PAIRED BOX PROTEIN PAX-3.  
 GN PAX3 OR PAX-3.  
 OS MUS MUSCULUS (MOUSE).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; RODENTIA.  
 RN  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE: 91216108.  
 RA GOULDING M.D.; CHALSPAKIS G., DEUTSCH U., ERSELIUS J.R., GRUSS P.;  
 RL EMBO J. 10:1135-1147(1991).  
 RN [2]  
 RP SEQUENCE OF 219-260 FROM N.A., AND VARIANT SPLITCH.  
 RX MEDLINE: 92034997.  
 RA EPSTEIN D.J.; VEKEMANS M., GROS P.;  
 RL CELL 67:767-774(1991).  
 RN [3]  
 RP VARIANT SPLITCH.  
 RX MEDLINE: 94010930.  
 RA VOGAN K.J., EPSTEIN D.J., TRASLER D.G., GROS P.;  
 RL GENOMICS 17:364-369(1993).  
 RL -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR ASSOCIATED WITH DEVELOPMENT OF ALVEOLAR RHABDOMYOSARCOMA.  
 -1- DEVELOPMENTAL STAGE: EXPRESSED DURING EARLY NEUROGENESIS.  
 -1- SUBCELLULAR LOCATION: NUCLEAR.  
 -1- SUBUNIT: CAN BIND AS A HETERODIMER WITH PAX7.  
 -1- DISEASE: THE SPLITCH (SP) MOUSE MUTANT DISPLAYS DEFECTS IN NEURAL TUBE CLOSURE IN THE FORM OF EXENCEPHALY AND SPINA BIFIDA.  
 -1- SIMILARITY: STRONG WITH OTHER "PAIRED-TYPE" HOMEOBOX PROTEINS.  
 -1- SIMILARITY: CONTAINS A PAIRED BOX DOMAIN.  
 DR EMBL: X59358; G53592; -;  
 DR EMBL: S66429; G239201; -;  
 DR EMBL: S66433; G239203; -;  
 DR PIR: S15031; S15031.  
 DR HSSP: P02836; 1ENH.  
 DR PROSITE: PS00027; HOMEOBOX.  
 DR PROSITE: PS00034; PAIRED\_BOX.  
 DR TRANSFAC; T00680; --

KW TRANSCRIPTION REGULATION; PAIRED BOX; DEVELOPMENTAL PROTEIN;  
 KW NUCLEAR PROTEIN; DNA-BINDING; HOMEOBOX; DISEASE MUTATION.  
 FT DOMAIN 34 159 PAIRED BOX.  
 FT DNA\_BIND 219 278 HOMEOBOX.  
 FT VARIANT 42 42 G -> R (IN SPLITCH).  
 SQ SEQUENCE 479 AA; 52984 MW; 985984B6 CRC32;

Query Match 19.9%; Score 369; DB 7; Length 479;

Best Local Similarity 66.7%; Pred. No. 3,62e-50;  
 Matches 46; Conservative 13; Mismatches 9; Indels 1; Gaps 1;

Db 215 lkrqrrttftaelelekaferthypdiyltreelaqrktlrearvqvwfsnriarw 274  
 :|||||:|||||: ||| | | :|||||:|||||: ||||||| |||||:  
 QY 29 LRRKQRRNRTTFALQQLALEAVFAQTHYDPVFTREELAMKINLTREARVQVWFQNRRAK 88

Db 275 rk-qagang 282

QY 89 RKTERGASD 97

RESULT 6  
 ID PAX3\_HUMAN STANDARD; PRT; 479 AA.  
 AC P23760;  
 DT 01-NOV-1991 (REL. 20, CREATED)  
 DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)  
 DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
 DE PAIRED BOX PROTEIN PAX-3 (HUP2).  
 GN PAX3.  
 OS HOMO SAPIENS (HUMAN).  
 OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
 OC EUTHERIA; PRIMATES.  
 RN [1]  
 RP SEQUENCE OF 1-29 AND 197-479 FROM N.A.  
 RA RIETHMAN H.C., MACINA R.A.;  
 RL SUBMITTED (JUL-1994) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [2]  
 RP SEQUENCE OF 30-195 FROM N.A.  
 RX MEDLINE: 89305521.  
 RA BURRI M., TROMVOUKIS Y., BOPP D., FRIGERIO G., NOLL M.;  
 RL EMBO J. 8:1183-1190(1989).  
 RN [3]  
 RP SEQUENCE OF 196-392 FROM N.A.  
 RX MEDLINE: 95072569.  
 RA TASSABEHJI M., NEWTON V., LEVERTON K., TURNBULL K., SEEMANOVA E.,  
 RA KUNZE J., SPERLING K., STRACHAN T., READ A.;  
 RL HUM. MOL. GENET. 3:1069-1074(1994).  
 RN [4]  
 RP SEQUENCE FROM N.A. (PAX3A AND PAX3B).  
 RX MEDLINE: 94171226.  
 RA TSUKAMOTO K., NAKAMURA Y., NIIKAWA N.;  
 RL HUM. GENET. 93:270-274(1994).  
 RN [5]  
 RP VARIANT WS1 ALA-63--ILE-67 DEL.  
 RX MEDLINE: 92168113.  
 RA TASSABEHJI M., READ A.P., NEWTON V.E., HARRIS R., BALLING R.,  
 RA GRUSS P., STRACHAN T.;  
 RL NATURE 355:635-636(1992).  
 RN [6]  
 RP VARIANT WS1 LEU-50.  
 RX MEDLINE: 92168114.  
 RA BALDWIN C.T., HOTH C.F., AMOS J.A., DA-SILVA E.O., MILUNSKY A.;  
 RL NATURE 355:637-638(1992).  
 RN [7]  
 RP VARIANT WS2 ALA-81.  
 RX MEDLINE: 93258399.  
 RA TASSABEHJI M., READ A.P., NEWTON V.E., PATTON M., GRUSS P., HARRIS R.,  
 RA STRACHAN T.;  
 RL NAT. GENET. 3:26-30(1993).  
 RN [8]  
 RP VARIANTS WS3 HIS-47 AND WS1 LEU-56.  
 RX MEDLINE: 93190976.  
 RA HOTH C.F., MILUNSKY A., LIPSKY N., SHEFFER R., CLARREN S.K.,  
 RA BALDWIN C.T.;  
 RL AM. J. HUM. GENET. 52:455-462(1993).  
 RN [9]  
 RP VARIANTS WS1 LEU-45 AND ASP-99.  
 RX MEDLINE: 95072569.  
 RA TASSABEHJI M., NEWTON V.E., LEVERTON K., TURNBULL K., SEEMANOVA E.,  
 RA KUNZE J., SPERLING K., STRACHAN T., READ A.P.;  
 RL HUM. MOL. GENET. 3:1069-1074(1994).  
 RN [10]

RP VARIANT WS1 VAL-62.  
RX MEDLINE: 95135456.  
RA PIERPONT J.W., DOOLAN L.D., AMANN K., SNEAD G.R., ERICKSON R.P.;  
RL HUM. MUTAT. 4:227-228(1994).  
[11]  
RN RP VARIANTS WS1 PHE-265 AND GLY-271.  
RX MEDLINE: 95126143.  
RA LALWANI A.K., BRISTER J.R., FEX J., GRUNDFAST K.M., PLOPLIS B.,  
RA SAN AGUSTIN T.B., WILCOX E.R.;  
RL AM. J. HUM. GENET. 56:75-83(1995).  
[12]  
RN RP VARIANT WS3 PHE-84.  
RX MEDLINE: 95243235.  
RA ZLOTOGORA J., LERER I., BAR-DAVID S., ERGAZ Z., ABELIOVICH D.;  
RL AM. J. HUM. GENET. 56:1173-1178(1995).  
[13]  
RN RP VARIANTS WS1 MET-60; GLU-85 AND SER-238.  
RX MEDLINE: 96042708.  
RA BALDWIN C.T., HOTH C.F., MACINA R.A., MILUNSKY A.;  
RL AM. J. MED. GENET. 58:115-122(1995).  
[14]  
RN RP VARIANT CDHS LYS-47.  
RX MEDLINE: 96263735.  
RA ASHER J.H. JR., SOMMER A., MORELL R., FRIEDMAN T.B.;  
RL HUM. MUTAT. 7:30-35(1996).  
[15]  
RN RP VARIANT LYS-315.  
RA HOL F.A., GEURDS M.P.A., CHATKUP T.S., SHUGART Y.Y., BALLING R.,  
RA SCHRANDE-STUMPEL C.T.R.M., JOHNSON W.G., HAMEL B.C.J.,  
RA MARIMAN E.C.M.; 33:655-660(1996).  
RL J. MED. GENET. 33:655-660(1996).  
CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR ASSOCIATED WITH  
DEVELOPMENT OF ALVEOLAR RHABDYOSARCOMA.  
CC -1- SUBCELLULAR LOCATION: NUCLEUS.  
CC -1- SUBUNIT: CAN BIND AS A HETEROIDIMER WITH PAX7.  
CC -1- ALTERNATIVE PRODUCTS: IN ADDITION TO THE NORMAL PAX3 PROTEIN, THE  
GENE CAN PRODUCE, BY ALTERNATIVE SPLICING, TWO TRUNCATED FORMS  
WHICH ARE CALLED PAX3A AND PAX3B.  
CC -1- DISEASE: DEFECT IN PAX3 ARE THE CAUSE OF WAARDENBURG'S SYNDROME  
(W5), AN AUTOSOMAL DOMINANT DISORDER, CHARACTERIZED WITH NON  
PROGRESSIVE SENSORY-NEURAL DEAFNESS, AS WELL AS PIGMENTARY CHANGES  
OF THE IRIDES AND OF THE HAIR AND SKIN; EACH OF THESE FEATURES  
MAY BE UNI- OR BILATERAL. ON THE BASIS OF THE PRESENCE OR ABSENCE  
OF DYSTOPIA CANTHORUM (LATERAL DISPLACEMENT OF THE INNER CORNER OF  
THE EYE), TYPE I (WS1) AND TYPE II (WS2): DOES NOT MAP TO PAX3) ARE  
DISTINGUISHED. TYPE III (WS3 OR KLEIN-WAARDENBURG SYNDROME) IS  
CHARACTERIZED WITH WS1 AND UPPER-LIMB INVOLVEMENT INCLUDING MUSCLE  
HYPOPLASIA WITH JOINT CONTRACTURES.  
CC -1- DISEASE: CRANIOFACIAL-DEAFNESS-HAND SYNDROME (CDHS), IS THOUGHT  
TO BE AN AUTOSOMAL DOMINANT DISEASE WHICH COMPRISES ABSENCE OR  
HYPOPLASIA OF THE NASAL BONES, HYPOPLASTIC MAXILLA, SMALL AND  
SHORT NOSE WITH THIN NARES, LIMITED MOVEMENT OF THE WRIST, SHORT  
PALMARAL FISSURES, ULNAR DEVIATION OF THE FINGERS, HYPERTELORISM  
AND PROFOUND SENSORY-NEURAL DEAFNESS.  
CC -1- DISEASE: RHABDYOSARCOMA-2 (RMS2) IS CHARACTERIZED BY A  
CHROMOSOMAL TRANSLOCATION T(2;13)(Q35;Q14) WHICH INVOLVES PAX3 AND  
FHR.  
CC -1- SIMILARITY: STRONG WITH OTHER "PAIRED-TYPE" HOMEBOX PROTEINS.  
CC -1- SIMILARITY: CONTAINS A PAIRED BOX DOMAIN.  
DR EMBL: U12263; G555818;  
DR EMBL: U12259; G555819;  
DR EMBL: U12258; G555819; JOINED.  
DR EMBL: U12260; G555819; JOINED.  
DR EMBL: U12262; G555819; JOINED.  
DR EMBL: X15043; E224055;  
DR EMBL: X15252; E224055; JOINED.  
DR EMBL: X15253; E224055; JOINED.  
DR EMBL: Z29972; -; NOT\_ANNOTATED\_CDS.  
DR EMBL: Z29973; -; NOT\_ANNOTATED\_CDS.  
DR EMBL: Z29974; -; NOT\_ANNOTATED\_CDS.  
DR EMBL: S69369; G545845;  
DR EMBL: S69370; G545847;  
PIR: S06960; S06960.

DR MIN: 193500; -;  
DR MIN: 148820; -;  
DR MIN: 268220; -;  
DR MIN: 122880; -;  
DR PROSITE; PS00027; HOMEBOX.  
DR PROSITE; PS00034; PAIRED\_BOX.  
DR TRANSFAC; T00397; -;  
DR TRANSCRIPTION; T00679; -;  
KW NUCLEAR PROTEIN; DISEASE MUTATION; CHROMOSOMAL TRANSLOCATION;  
KW PROTO-ONCOGENE; ALTERNATIVE SPLICING; POLYMORPHISM.  
FT DOMAIN 34 159  
FT DNA\_BIND 219 278  
FT VARSPLIC 197 215  
FT VARSPLIC 216 479  
FT VARSPLIC 197 206  
FT VARSPLIC 207 479  
FT VARSPLIC 45 45  
FT VARSPLIC 47 47  
FT VARSPLIC 47 47  
FT VARSPLIC 50 50  
FT VARSPLIC 56 56  
FT VARSPLIC 60 60  
FT VARSPLIC 62 62  
FT VARSPLIC 81 81  
FT VARSPLIC 84 84  
FT VARSPLIC 85 85  
FT VARSPLIC 99 99  
FT VARSPLIC 238 238  
FT VARSPLIC 265 265  
FT VARSPLIC 271 271  
FT VARSPLIC 315 315  
SQ SEQUENCE 479 AA; 52967 MW; 6EE1491D CRC32;  
Query Match 19.9%; Score 369; DB 7; Length 479;  
Best Local Similarity 66.7%; Pred. No. 3.62e-50;  
Matches 46; Conservative 13; Mismatches 9; Indels 1; Gaps 1;  
Db 215 lkrqtrrttftaeqlleerferthypdytreelaqrakltearvqvfnrtarv 274  
Qy 29 LRRQRNRTTFAQLQLEAVEPAQTHYDVTREELAMKINLTARVQVFNQRRAK 88  
Db 275 rk-qagang 282  
Qy 89 RKTERGASD 97  
RESULT 7  
ID PAX7 MOUSE STANDARD; PRT; 290 AA.  
AC P47239;  
DT 01-FEB-1996 (REL. 33, CREATED)  
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
DE PAIRED BOX PROTEIN PAX-7 (FRAGMENT).  
GN PAX7 OR PAX-7.  
OS MUS MUSCULUS (MOUSE).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; RODENTIA.  
RN [1]  
RX MEDLINE: 91265334.  
RA JOSTES B., WALTHER C., GRUSS P.;  
RL MECH. DEV. 33:27-37(1990).  
CC -1- FUNCTION: PROBABLE TRANSCRIPTION FACTOR. IT MAY HAVE A ROLE IN  
MYOGENESIS.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- SUBUNIT: CAN BIND AS A HETEROIDIMER WITH PAX3.  
CC -1- DEVELOPMENTAL STAGE: EXPRESSED SPECIFICALLY DURING THE DEVELOPMENT  
OF THE NERVOUS AND MUSCULAR SYSTEM.  
CC -1- SIMILARITY: STRONG WITH OTHER "PAIRED-TYPE" HOMEBOX PROTEINS.

CC -1- SIMILARITY: CONTAINS A PAIRED BOX DOMAIN.  
DR EMBL; 020792; G736381; -  
KW TRANSCRIPTION REGULATION; HOMEBOX; DNA-BINDING; NUCLEAR PROTEIN;  
KW DEVELOPMENTAL PROTEIN; PAIRED BOX.  
FT NON\_TER 1  
FT DOMAIN <1 125 PAIRED BOX.  
FT DNA\_BIND 181 240 HOMEBOX.  
FT NON\_TER 230  
SQ SEQUENCE 290 AA; 32911 MW; 51176DAF CRC32;  
Query Match 19.4%; Score 360; DB 7; Length 290;  
Best Local Similarity 65.2%; Pred. No. 2.42e-48;  
Matches 45; Conservative 13; Mismatches 10; Indels 1; Gaps 1;  
Db 177 lkrqrrstttfaeqllealekaferthypdytreelagrtkltearfqvwsnrarw 236  
Y 29 LRRKQRRNRTTFALQOLEALEAVFAQTHYPDVFTREELAMKINLTARVQVWFQNRRAKW 88  
Db 237 rk-qagang 244  
Y 89 RKTERGASD 97  
RESULT 8  
ID GSC\_DRONE STANDARD; PRT; 419 AA.  
AC P54366;  
DT 01-OCT-1996 (REL. 34, CREATED)  
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)  
DT 01-OCT-1996 (REL. 34, LAST ANNOTATION UPDATE)  
DE HOMEBOX PROTEIN GOOSECOID.  
GN GSC.  
OS DROSOPHILA MELANGASTER (FRUIT FLY).  
OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 96202483.  
RA CORLEY A., STELLA M., COFFINIER C., KESSLER D., MAILLOS C.,  
RA DESAIN S., DESPLAN C.;  
RL DEVELOPMENT 122:1641-1650(1996).  
CC -1- SIMILARITY: STRONG WITH OTHER "PAIRED-TYPE" HOMEBOX PROTEINS.  
DR EMBL; X95420; E220356; -  
DR FLYBASE; FBGN007777; GSC.  
KW DEVELOPMENTAL PROTEIN; NUCLEAR PROTEIN; DNA-BINDING; HOMEBOX.  
FT DOMAIN 104 107 POLY-ALA.  
FT DOMAIN 164 169 POLY-SER.  
FT DOMAIN 195 199 POLY-ALA.  
FT DNA\_BIND 286 345 HOMEBOX.  
SQ SEQUENCE 419 AA; 44949 MW; D6880216 CRC32;  
Query Match 19.2%; Score 357; DB 4; Length 419;  
Best Local Similarity 49.5%; Pred. No. 9.79e-48;  
Matches 45; Conservative 16; Mismatches 29; Indels 1; Gaps 1;  
Db 262 phl-gahhghqhlshlbgpppkrrhrtrfteeqleatfdkthypdvrlreqla 320  
Y 8 POLETAPEFGNHSITGDDGFLRRKQRRNRTTFALQOLEALEAVFAQTHYPDVFTREELA 67  
Db 321 lkvdikeervvfkrrakrkqkeeqer 351  
Y 68 MKINLTARVQVWFQNRRAKRWKTERGASDQ 98  
RESULT 9  
ID PAX5\_CHICK STANDARD; PRT; 216 AA.  
AC P47237;  
DT 01-FEB-1996 (REL. 33, CREATED)  
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
DE PAIRED BOX PROTEIN PAX-5 (FRAGMENT).  
GN PAX-6.  
OS GALLUS GALLUS (CHICKEN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;

OC GALLIFORMES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 94170909.  
RA LI H.S., YANG J.M., JACOBSON R.D., PASKO D., SUNDIN O.;  
RL DEV. BIOL. 162:181-194(1994).  
CC -1- FUNCTION: MAY BE A TRANSCRIPTION FACTOR WITH IMPORTANT FUNCTIONS  
CC IN EYE AND NASAL DEVELOPMENT.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- SIMILARITY: CONTAINS A PAIRED BOX DOMAIN.  
DR EMBL; S69508; G545839; -  
KW TRANSCRIPTION REGULATION; HOMEBOX; DNA-BINDING; NUCLEAR PROTEIN;  
KW DEVELOPMENTAL PROTEIN; PAIRED BOX.  
FT NON\_TER 1  
FT DNA\_BIND 10 69 HOMEBOX.  
SQ SEQUENCE 216 AA; 24469 MW; DAB2288B CRC32;  
Query Match 19.1%; Score 354; DB 7; Length 216;  
Best Local Similarity 57.1%; Pred. No. 3.95e-47;  
Matches 44; Conservative 18; Mismatches 15; Indels 0; Gaps 0;  
Db 6 lkrklqrntstqlealekeferthypdvfarerlaakidlpairlqvwsnrakw 65  
Y 29 LRRKQRRNRTTFALQOLEALEAVFAQTHYPDVFTREELAMKINLTARVQVWFQNRRAKW 88  
Db 66 reeklrnqrqasntp 82  
Y 89 RKTERGASDQEPGAKEP 105  
\*RESULT 10  
ID PAX6\_COTJA STANDARD; PRT; 416 AA.  
AC P47238;  
DT 01-FEB-1996 (REL. 33, CREATED)  
DT 01-FEB-1996 (REL. 33, LAST SEQUENCE UPDATE)  
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)  
DE PAIRED BOX PROTEIN PAX-6 (PAX-QNR).  
GN PAX-6.  
OS COTURNIX COTURNIX JAPONICA (JAPANESE QUAIL).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; AVES; NEOGNATHAE;  
OC GALLIFORMES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX TISSUE-NEURORETINA;  
RX MEDLINE; 92366156.  
RA MARTIN P., CARRIERE C., DOZIER C., QUATANNENS B., MIRABEL M.A.,  
RA VANDERBINDER B., STEHELIN D., SAULE S.;  
RL ONCOGENE 7:1721-1728(1992).  
CC -1- FUNCTION: MAY BE A TRANSCRIPTION FACTOR WITH IMPORTANT FUNCTIONS  
CC IN EYE AND NASAL DEVELOPMENT.  
CC -1- SUBCELLULAR LOCATION: NUCLEAR.  
CC -1- SIMILARITY: STRONG WITH OTHER "PAIRED-TYPE" HOMEBOX PROTEINS.  
CC -1- SIMILARITY: CONTAINS A PAIRED BOX DOMAIN.  
DR EMBL; X70475; G311772; -  
KW TRANSCRIPTION REGULATION; HOMEBOX; DNA-BINDING; NUCLEAR PROTEIN;  
KW DEVELOPMENTAL PROTEIN; PAIRED BOX.  
FT DOMAIN 4 128 PAIRED BOX.  
FT DOMAIN 129 209 GLN/GLY-RICH.  
FT DNA\_BIND 210 269 HOMEBOX.  
SQ SEQUENCE 416 AA; 46027 MW; D636633C CRC32;  
Query Match 19.1%; Score 354; DB 7; Length 416;  
Best Local Similarity 57.1%; Pred. No. 3.95e-47;  
Matches 44; Conservative 18; Mismatches 15; Indels 0; Gaps 0;  
Db 206 lkrklqrntstqlealekeferthypdvfarerlaakidlpairlqvwsnrakw 265  
Y 29 LRRKQRRNRTTFALQOLEALEAVFAQTHYPDVFTREELAMKINLTARVQVWFQNRRAKW 88  
Db 266 reeklrnqrqasntp 282  
Y 89 RKTERGASDQEPGAKEP 105

3



```

QY      89 RKTER 93

RESULT 14
ID      HMPR_DROME      STANDARD;      PRT;      613 AA.
AC      P06601;
DT      01-JAN-1988 (REL. 06, CREATED)
DT      01-JAN-1988 (REL. 06, LAST SEQUENCE UPDATE)
DT      01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE      SEGMENTATION PROTEIN PAIRED.
PRD.
OS      DROSOPHILA MELANOGASTER (FRUIT FLY).
OC      EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE; 87051745.
RA      FRIGERIO G., BURRI M., BOPP D., BAUMGARTNER S., NOLL M.;
RL      CELL 47-735-746(1986).
CC      -1- FUNCTION: PRD IS A PAIR-RULE PROTEIN REQUIRED FOR SEGMENTATION IN
CC      DROSOPHILA. CAPABLE OF SEQUENCE-SPECIFIC DNA-BINDING.
CC      -1- SUBCELLULAR LOCATION: NUCLEAR.
CC      -1- SIMILARITY: STRONG WITH OTHER "PAIRED-TYPE" HOMEOBOX PROTEINS.
CC      -1- SIMILARITY: CONTAINS A PAIRED BOX DOMAIN.
DR      EMBL; M14548; G158160;
DR      FIR; A26062; A26062.
DR      HSSP; P02836; LENH.
DR      FLYBASE; FBGN0003145; PRD.
DR      PROSITE; PS00027; HOMEOBOX.
DR      PROSITE; PS00034; PAIRED_BOX.
DR      TRANSFAC; T00699;
KW      HOMEOBOX; DNA-BINDING; DEVELOPMENTAL PROTEIN; NUCLEAR PROTEIN;
KW      PAIRED BOX; REPEAT; PAIR-RULE PROTEIN; TRANSCRIPTION REGULATION.
FT      DOMAIN 27 151 PAIRED BOX.
FT      DNA_BIND 213 272 HOMEOBOX.
FT      DOMAIN 552 572 10 X 2 AA TANDEM REPEATS OF H-P.
FT      VARIANT 164 164 A -> T.
FT      VARIANT 220 220 F -> I.
SQ      SEQUENCE 613 AA; 65497 MW; 0922346E CRC32;

Query Match 19.0%; Score 353; DB 5; Length 613;
Best Local Similarity 47.1%; Pred. No. 6-29e-47;
Matches 49; Conservative 26; Mismatches 26; Indels 3; Gaps 3;

Db 202 esepgialkrkqrcttfsaqldeleafertqydydiyreelaqrntleairqvfwf 261
: : : : : ||||| ||||| : : : : : ||||| ||||| : : ||||| |||||
QY 23 DFDGDF-LRRKQRNRRTTALQOLEAVEFAQTHYPDVFTREELAMKINLTARVQVWF 81
262 snrrarlkrqhtsvsgagpggaasvshvaassslpsvsvssvps 305
||||| : : : : : ||||| : : : : : ||||| : : : : : || : :
QY 82 QNRRAKWRKTERGASDQEPG-AKEPMAEVTTPP-VRNINSPPEG 123

RESULT 15
ID      GSBP_DROME      STANDARD;      PRT;      449 AA.
AC      P09083;
DT      01-NOV-1988 (REL. 09, CREATED)
DT      01-NOV-1988 (REL. 09, LAST SEQUENCE UPDATE)
DT      01-FEB-1995 (REL. 31, LAST ANNOTATION UPDATE)
DE      GOOSEBERRY PROXIMAL PROTEIN (BSH4).
GN      GSB-P OR GSBA.
OS      DROSOPHILA MELANOGASTER (FRUIT FLY).
OC      EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE; 88112802.
RA      BAUMGARTNER S., BOPP D., BURRI M., NOLL M.;
RL      GENES DEV. 1:1247-1267(1987).
RN      [2]
RP      SEQUENCE OF 12-145 AND 162-241 FROM N.A.
RX      MEDLINE; 87051758.
RA      BOPP D., BURRI M., BAUMGARTNER S., FRIGERIO G., NOLL M.;
RL      CELL 47:1033-1040(1986).
CC      -1- FUNCTION: GOOSEBERRY DISTAL AND PROXIMAL ARE SEGMENT-POLARITY

```